Dynamicity in Social Trends towards Trajectory Based Location Recommendation

Muhammad Aamir Saleem, Young-Koo Lee, Sungyoung Lee
Department of Computer Engineering, Kyung Hee University, South Korea
{aamir,yklee,sylee}@oslab.khu.ac.kr

Abstract. The pervasiveness of location acquisition technologies has significantly elevated the demands of experience sharing recommendation systems. These systems are highly affected by social dynamicity and trends which are not exploited in existing studies. In this paper, we have proposed a GPS trajectory focused approach that endorses interesting locations. Tree based hierarchical clusters of visited locations are utilized to incorporate the timely changing social trends and personalized preferences of the users. Experimental studies are conducted on real world dataset for verification and validity of the proposed technique.

Keywords: Trend Analysis, GPS trajectories, Location recommendation.

1 Introduction

In recent years, everywhere in the world, the trend of using locations attainment technologies such as GPS, GSM, and Wi-Fi have grown up rapidly in different domains. Especially the GPS based path finder devices in transport systems has been proved a competent way to extract the traces of moving objects. These traces are being used to exploit the human behavior and social dynamicity [1]. Recently, a huge number of social networks and websites have also enabled people to establish Geo-related web communities, and share their daily life experiences. Experiences related to movements such as visiting locations, movement paths etc. help an individual to learn about an unknown area in a short period with minimal efforts. Meanwhile, such information also helps mobile guides and recommendation systems [2-3].

However, deducing the interest of a user and the recommendation of an appropriate location is quite complex phenomenon. Since it depends on a number of factors e.g. location history of users and their corresponding experiences, ease of attainment and social importance of the location. Another important measure, which is being ignored by existing techniques, is the effect of dynamically changing social trends on evolution of user's interest. A recent study [6] show that change in social trends has impacted the renovation of interests of people, enormously. Since peoples' experiences and social dynamicity lead the world towards easy and better solutions. Therefore people want to transform their routine activities by overwhelming these trends. For example, consider the case of a diabetic patient who used to visit a physician at clinic X, for his weekly checkup but he recently started visiting clinic Y because of the possible bad experience and displeasure he got. Thus, his current location history shows the preference of clinic Y, however historic patterns express the betterment of clinic X. Considering the above scenario existing techniques recommend clinic X to the user due to its higher frequency. However, clinic Y should be recommended due to the change in preference of the user based on his experience of historic adoptions.

In this paper, we aim to mine top-k interesting locations by using GPS trajectories of multiple users. Assimilation of correlation between the user and location, and timely changing trends are included to reflect state-of-the-art interest of people. In step 1, the visited location histories of the users are identified by using GPS traces and in second step it is divided into N time intervals to incorporate the effect of changing interest. In step 3, for each time interval a hierarchical clustering based tree is constructed. The clustering in accomplished based on the nature of the visit (performed activity) and region of location. In step 4, evaluation of each location is done based on both of the peripheral and personalized stimuli. Peripheral stimuli include social importance and experience of the visitors of the location. However, personalized stimuli incorporate the user's personalized preferences e.g. time and cost of reaching that location. Finally, skyline computation technique, LSA is applied to fetch the most preferred interesting location according to both stimuli.

The summary of main contributions of the paper includes the introduction of a novel location recommendation measure based on dynamically changing social trends, evaluation of both peripheral and personalized stimuli of the location and experimental assessment of the proposed technique on real data sets. The rest of the paper is organized as follows. Section 2 describes related work. The preliminaries are discussed in section 3. Computations of location rank score and working of Skyline location recommendation are shown in section 4 and 5 respectively. Section 6 exploits the results of experiments and conclusion is given in section 7.

2 Related Work

Recommendation systems [2-3] use the public experiences to provide help in exploring the community. They effectively identify the content of interest from a potentially overwhelming set of choices. Collaborative filtering and nearest neighborhood approaches [9] are exploited to find the use of similar activities between users, and rank the options based on their correlation. Further, Pearson correlation and cosine similarity [10] are also widely used correlation measures. Another category of run-time mobile guide and recommendation systems [3] [11] estimates the individual preferences based on location history and target the recommendations. Moon et al, in [11] proposed the map-based personalized recommendation system. In this study, preferences of user are modeled by Bayesian network and history data is exploited to fetch most potential recommendation. However, in our proposed approach a novel measure of dynamically changing trends of the interest of domain users is presented to find their correlation. Second, we combine both the real world social impact of places and user's personalized measure to reflect their importance towards categorization of available choices.

3 Preliminaries

Definition 1: Imperative Location (L_I): An imperative location L_I is a group of trajectory points at a geographic region where a user stayed over a certain interval of time i.e. $L_I = \{p'_m, p'_{m+1}, p'_{m+2} \ldots p'_k\}$. The extraction of a stay point depends on three scale parameters, a time threshold T_{min}, a minimum and maximum distance threshold D_{min}, D_{max} and speed of a moving object s_{min}. These thresholds ensure the importance of location based on time, distance and speed [3]. So, a location will be considered imperative only if all of these three thresholds are satisfied i.e. $T_S \geq T_{min}$, $D_{max} \leq D_S \geq D_{min}$ and $s \leq s_{min}$; Where, T_S is the total time spent at a particular place, D_S is the distance area of a location and s is the speed of a moving object. Haversine formula [8] is used to find the distance and speed between two GPS points. Identification of imperative location is triggered on sequence of fulfillment of s_{min}, D_{min}, D_{max} and T_{min} respectively.

Definition 2: Location History (H_{Loc}): A location history is a set of visited imperative locations of a particular user i.e., $H_{Loc} = \{(L_{I_1}, t'_1), (L_{I_1}, t'_2), \ldots, (L_{I_1}, t'_n))\}$. It is further sub-divided into N number of timestamps T, based on t'_i where a threshold time T_t, represents time stamp T_i, where $t_j \leq T_t \leq t_k$.

Definition 3: Tree based Hierarchical Clustering *(TBHC)*: A hierarchical clustering based tree is constructed for location history of each time stamp T_i. Clustering of imperative locations is accomplished on the basis of nature and region of imperative locations. However, nodes of *TBHC* are the nature based clusters of imperative locations called Activity Cluster and region based clustering is consummate with in activity clusters.

Definition 4: Activity Cluster ($C_{i'j'}$): Activity cluster is the set of imperative locations of same activity category at level *i* of *TBHC*. Each child of $C_{i'j'}$ contributes the lower granularity level of activity of locations as compared to its parent i.e. $C_{i'} = \{C_{i'1}, C_{i'2} \ldots C_{i'm}\}$, Where $C_{i'j'}$ is a set of imperative locations of same nature, at more granular level than $C_{(i'-1)j'}$. For example consider a node of *TBHC*, an activity cluster $C_{11} \in$ *Clinics in Seoul (city)*. Children of this C_{11} will be sub-divided based on nature of clinics, e.g. Clinic for diabetic C_{21} and cancer patients C_{22}, will be clustered as two separate children of C_{11}, respectively (Shown in figure 1).

Definition 5: Location Cluster ($C_{i'j'k'}$): Location Cluster divides the activity cluster based on regions of imperative locations. Thus $C_{i'j'k'}$ is the set of imperative locations of same category in a particular region. i.e., $C_{i'j'k'} = \{C_{i'j'1}, C_{i'j'2} \ldots C_{i'j'l}\}$. In figure 1, two blue color sub clusters of C_{21} are showing the diabetes clinic in southern and northern part of Seoul respectively.

Problem Statement: Recommend the top-k interesting locations to the user for potential visits based on dynamically changing peripheral (external factors) and personalized (user's personalized) stimuli.

4 Computation of Location Rank Score *(S_{LR})*

Each L_I is evaluated to find a score called S_{LR}. Recommendations of locations are based on these scores. The detail computation of S_{LR} is given below.

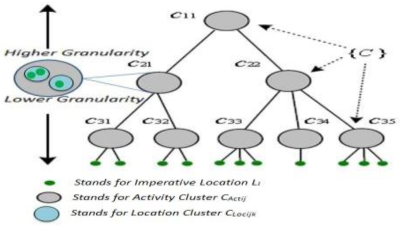

Fig.1.Tree based Hierarchical Clustering

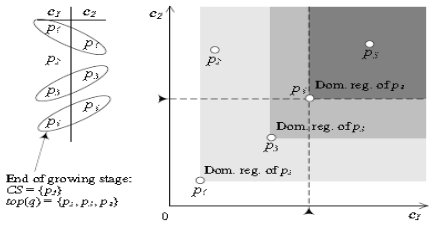

Fig. 2. LSA: Top k skyline facilities[9]

Since H_{Loc} of various people are uneven and incomparable thus the stay points pertaining to different individuals are not identical. *TBHC* is proposed to model the location history of multiple users. In this technique [4] is modified to incorporate the two level clustering at each step of the tree and to express the relationship of both nature and region cluster nodes at different granularities. *TBHC* is a hierarchical tree of *Activity clusters*, shown in gray color in the Figure 1. Each node of the tree represents the category of set of similar L_I of activities, (shown in green color). These nodes are further drilled into children based on category division of imperative locations (Definition 3, 4, 5). To identify the significance of a location at a particular area, each of the *Activity Cluster* nodes is further clustered on the basis of region of corresponding L_I. Here point to be noted is that while moving from parents to children granularity of both of location and activity cluster changes i.e. granularity of nature of activity clusters, and region of location clusters of imperative locations increases at each step of *TBHC*. To fetch the interesting locations for a particular user, activity and location clusters of all the time stamps of all the users (friends) are examined and stimuli are calculate based on peripheral stimuli and personalized stimuli. Computation of both of the measures is given below in detail.

4.1 Peripheral Stimuli (*S_p*) Computation

The interest of a location is highly influenced by the measure of its importance in society. S_P incorporates the effect of external factors of society on an L_I. These external factors include significance of visiting location, and visitors' experiences. S_p is calculated by a monotonic function, given by following equation.

$$s_p = \alpha(v_{fL_I}) + \beta(u_s) \tag{1}$$

Where;

s_p = Influence scored of external factors (peripheral stimuli)

v_{fL_I} = Number of visit of L_I at *i-th* level of TBHC

u_s =Visitor experience of L_I

α, β = personalized coefficients to assign the weightage of visiting frequency and user experience respectively.

Usually popular places are visited more frequently thus we consider visiting frequency as directly proportional to the S_P. Here an important point is that numbers of visit in latest timestamp are assigned higher score as compared to previous one, to consider the dynamically changing social trend. It is given by following equation.

$$v_f = \sum_{T=1}^{N} \frac{T}{N} \left(\frac{v_{fUL_I} C_{i'j'k'}}{\sum_{j=1}^{m} \sum_{k=1}^{l} v_{fU} C_{i'jk}} \right) \tag{2}$$

Where;

$v_{fUL_I} C_{i'j'k'}$	= Number of visit of location L_I by all users at location clusters $C_{i'j'k'}$
$\sum_{j=1}^{m} \sum_{k=1}^{l} v_{fU} C_{i'jk}$	= Number of visits of all users at all the locations clusters of all the activity clusters at level *i*
T, N	= Time stamp and total number of time stamps of *TBHC* , respectively

The visitor's experience is also considered as an important parameter to evaluate a location. The overall historic experience of a visitor towards locations in a particular region and also to corresponding category is evaluated for each of time stamp T_i. After assigning the higher weight to timestamp of recently visited imperative locations, an overall rank score of a visitor u_s is computed. The two factors involved in computation of u_s are given below in the equation 3. First part of equation (fractions in first parenthesis) shows the importance of location for a visitor *u'*, among his visited location history and second (fractions in second and last parenthesis) is for assessment of the importance of visitor *u'* among all the visitors *U*. It is given by equation 3.

$$u_s = \sum_{T=1}^{N} \frac{T}{N} \{ (\frac{v_{fu'L_I} C_{i'j'k'}}{v_{fu'} C_{i'j'k'}} + \frac{v_{fu'} C_{i'j'k'}}{\sum_{k=1}^{l} v_{fu'} C_{i'j'k}} + \frac{\sum_{k=1}^{l} v_{fu'} C_{i'j'k}}{\sum_{j=1}^{m} \sum_{k=1}^{l} v_{fu'} C_{i'jk}}) + (\frac{\sum_{j=1}^{m} \sum_{k=1}^{l} v_{fu'} C_{i'jk}}{\sum_{j=1}^{m} \sum_{k=1}^{l} v_{fU} C_{i'jk}}) \} \tag{3}$$

4.2 Personalized Stimuli (*S_{prn}*) Computation

User's personalized preferences toward any location also hold significant importance towards selection of it as a potential recommended position [8] [13]. Since usually visitor incorporates their own likings and interest, while planning to visit some location such as distance, cost and time of reaching at location. Thus in this section we incorporate these preferences of user (who request/query for location recommendation) called Personalized Stimuli S_{prn}. S_{prn} is the set of user's personalized preferences and their assigned preference weights i.e. S_{prn} ={(p_1,w_1), (p_2,w_2),… (p_m,w_m)},where p_i is the individual preference of user *u'* and w_i is its corresponding weight, for selection of interesting location L_{In}. All of these preferences S_{prn} and monotonic function score of peripheral stimuli S_p is send to Skyline location recommender for ultimate evaluation.

5. Skyline Location Recommendation

Each imperative location is evaluated, based on multi-cost parameters i.e. S_p and S_{prn}. Thus to assign corresponding importance to each of parameter we used skyline computation technique (*LSA*) [5]. *LSA* finds the entity that is best, based on all the concerned measures. Thus the purpose of using it is to pin the location that is most preferred among all the evaluated parameters. A general concept and working of *LSA* is shown in figure 2.

Score of peripheral stimuli is considered as a single parameter because it covers the evaluation of the social impact of location. While S_{prn} is the set of personalized presences and each of it is deliberated as separate individual parameter i.e. $S_{prn} = \{(p_1, w_1), (p_2, w_2), (p_3, w_3)\}$, where the p_1 is cost, p_2 is time and p_3 is distance parameter of a user u'. After evaluation on the basis of their corresponding weights, and provision of S_p for each imperative location, *LSA* compute the *top-k* interesting locations that are best based on each of these four parameters i.e. cost, time, distance and peripheral stimuli S_p.

6. Results

For the experimental evaluation, we used the dataset of Geo life trajectory [4] [7] to evaluate our proposed approach. It contains 17,621 trajectories with corresponding transportation mode. We visualize this dataset in a sense that transportation modes of trajectories are considered as imperative locations and use of transportation mode is deliberated as a visit of imperative location. Transport modes are considered as nature of imperative location. Thus *Activity Clusters* include the users, having same transport mode, and number of using of same transport portrays the frequency of visiting an imperative location in a *Category Cluster*. For *Location Cluster*, we divide the dataset into 11 groups based on region of observed trajectory. Further data set is divided into 6 equal time intervals of 6 months each, to incorporate the dynamicity and changing of social trend. Each time interval represents the time stamp T_i of TBHC. Thus, here we are recommending the *top-k* transportation modes by incorporating the changing social trends to depict the location recommendation.

TBHC is built for each T_i of all the 11 groups. We include the results of group 1 due to maximum availability of data for all intervals, except interval 6 which is ignored in the results. Fig.3 represents the normalized weight of using transport modes by all the users for corresponding time stamp T_i. The trend of using transportation mode is shown in the Fig. 4. The absolute stimuli shows the total frequency base result but trend analysis stimuli demonstrates the peripheral stimuli based on changing trends, given by equation 1(α, β are assigned equal weight, '1' to ignore dominance on each other). It can be observed that result of trend analysis stimuli is different than absolute one. In absolute measure *top-3* transportation modes are taxi, bike and bus respectively. However, result of trend analysis shows bus, taxi and bike as *top-3* interesting transportation mode respectively.

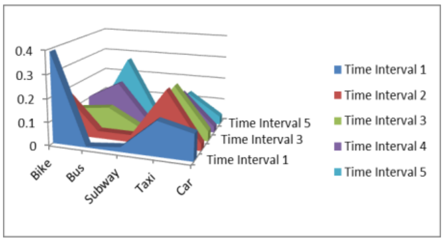

Fig. 3. The trend of using transportation modes over different time intervals

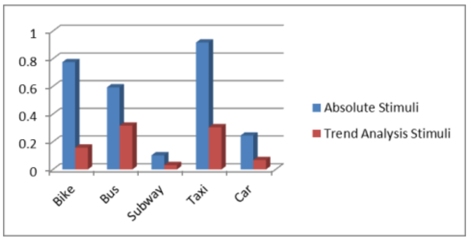

Fig. 4. Changing trend of Transportation modes

To calculate the peripheral stimuli we assumed three preference parameters, exercise, time and cost of using transport. Preference weight to each of these parameters (p_i) is assigned randomly e.g., (exercise, 0.3), (time, 0.5) and (cost, 0.2). Further each of transportation modes is evaluated based on these parameters and weights are assigned intuitively. Evaluation of transportation and preference score for each of p_i is shown in Table 1. Further, *LSA* is applied on it and pinning of transportation modes as skyline facility is shown by different colors. The representation of each color to corresponding transport mode is shown in the last two columns. The transportation mode that is pinned firstly by all the parameters (columns) is considered as most potential mode for recommendation e.g. Bus in the table 1. Similarly all the transportation modes are evaluated and are shown in last column of table 1, in descending order of their priority.

Table 1: Top-K interesting modes by LSA, based on both S_p and S_{prn}

Exercise Stimuli	Time stimuli	Cost stimuli	Peripheral stimuli		
1.8	0.5	0.4	0.317		=Bus
0.9	1.0	0.6	0.305		=Car
0.9	1.5	0.6	0.158		=Bike
0.6	1.5	1.4	0.068		=Taxi
0.3	4.0	1.8	0.032		=Subway

7. Conclusion

Social trends that change dynamically put a significant effect on renovation of people interests. In this paper, we explored this phenomenon as a vital measure for real life experience sharing based recommendation systems. The proposed approach incorporates both peripheral and personalized stimuli with changing social trends to categorize the interesting places for possible recommendations. Experimental evaluation on dataset of real life trajectories of users is performed to verify and validate the method. Results show that interest of people change over the interval of time and combining both peripheral and personalized stimuli on the basis of these changing interests, substantially affect the selection criteria of recommendation approaches. Author considers treatment sharing social networks such as patientlikeme, location based social networks and experience based recommender systems such as emerging shopping malls and best physician clinics for a particular disease, as potential applications of the proposed approach.

Acknowledgments. This work was supported from the NIPA (National IT Industry Promotion Agency), Global IT Talents Program in 2013.

References

1. M. A. Saleem, I. Fatima, K. U. Khan, Y.K. Lee and S. Y.Lee, "Trajectory Based Activity Monitoring and Healthcare Provisioning", The Tenth IEEE International Conference on Pervasive Intelligence and Computing (PiCom2012), Changzhou, China, December, 2012
2. Beeharee, Ashweeni and Steed "Exploiting real world knowledge in ubiquitous applications", in the proceeding of Personal and Ubiquitous Computing, 2007.
3. Mamoulis, N. et al. Mining, Indexing and Querying Historical Spatiotemporal Data. In Proceedings of KDD'04, USA, August 2004
4. Y. Zheng, L. X. Xie, W. Ma, Mining Interesting Locations and Travel Sequences from GPS Trajectories, In the proceeding of WWW, Madrid, Spain, 2009.
5. k. Mouratidis, Y. Lin, M. L. Yiun, "Preference Queries in Large Multi-Cost Transportation Network", in the proceeding of ICDE 2010.
6. B. Bishop, "Four voices on the future of food shopping http://www.brickmeetsclick.com/four-voices-on-the-future-of-food-shopping-1 , March 12, 2012.
7. Yu Zheng, Xing Xie, Wei-Ying Ma, GeoLife: "A Collaborative Social Networking Service among User, location and trajectory", In IEEE Data Engineering Bulletin. 33, 2, 2010.
8. Doctor Rick, Deriving the Haversine Formula, The Math Forum, April, 1999 http://www.movable-type.co.uk/scripts/latlong.html
9. Jarvelin, K., Kekalainen, J. Cumulated gain-based evaluation of IR techniques, ACM Transactions on Information Systems, 2002
10. Spertus, E., Sahami, M., Buyukkokten, Orkut., Evaluating similarity measures: a large-scale study in the orkut social network, In Proc. of KDD'05, 2005
11. Moon et all, "Location Based Recommendation systems Using Bayesian User's preference Model in Mobile Devices", in proceeding of LNCS, UIC , 2007.